RESEARCH REVIEW JOURNALS
Volume-04
Issue-03
March -2019

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

A Study of Distributes and Multi Agent System Towards Data Mining Approach

[1]Istiyaque Ahmad & [2]Dr. Yashpal

[1]*Research Scholar, OPJS University, Churu Rajasthan (India)*
[2]*Associate Professor, OPJS University, Churu Rajasthan (India)*

ARTICLE DETAILS

Article History
Published Online: 13 March 2019

Keywords
Distributed Data Mining, Multi-Agent Systems, Multi Agent Data Mining, Agent Based Distributed Data Mining.

ABSTRACT

The Data Mining technology normally adopts data integration method to generate Data warehouse, on which to gather all data into a central site, and then run an algorithm against that data to extract the useful Module Prediction and knowledge evaluation. However, a single data-mining technique has not been proven appropriate for every domain and data set. Data mining techniques involving in such complex environment must encounter great dynamics due to changes in the system can affect the overall performance of the system. Agent computing whose aim is to deal with complex systems has revealed opportunities to improve distributed data mining systems in a number of ways. Multi-agent systems (MAS) often deal with complex applications that require distributed problem solving. The field of Distributed Data Mining (DDM) deals with these challenges in analyzing distributed data and offers many algorithmic solutions to perform different data analysis and mining operations in a fundamentally distributed manner that pays careful attention to the resource constraints. Since multi-agent systems are often distributed and agents have proactive and reactive features which are very useful for Knowledge.

1. Introduction

Originated from knowledge discovery from databases (KDD), also known as data Mining (DM), distributed data mining (DDM) mines data sources regardless of their physical locations. The need for such characteristic arises from the fact that data produced locally at each site may not often be transferred across the network due to the Excessive amount of data and privacy issues. Recently, DDM has become a critical component of knowledge-based systems because its decentralized architecture reaches every networked business. Data Mining still poses many challenges to the research community. The main challenges in data mining are:

- Data mining to deal with huge amounts of data located at different sites The amount of data can easily exceed the terabyte limit;
- Data mining is very computationally intensive process involving very large data sets. Usually, it is necessary to partition and distribute the data for parallel processing to achieve acceptable time and space performance;
- Input data change rapidly. In many application domain data to be mined either is produced with high rate or they actually come in streams. In those cases, knowledge has to be mined fast and efficiently in order to be usable and updated;

Since the data comes from a diverse range of sources, including social networking sites, supply chains and government databases; it is usually unstructured following no particular format or layout. Hence, we need to process the incoming data to find useful information in it. Data mining is the process used wherein intelligent methods to extract interesting data patterns and knowledge from large amounts of data [1]. However, the rate at which data is produced is very high, and we need efficient methods of mining to keep abreast with it. To ensure higher performance, we use the concept of agents to support the data mining process known as agent mining. The distributed nature of agent mining brings several advantages to data mining such as autonomy, scalability, reliability, security, interactivity and high speed [2]. Agents can be used to automate the various tasks like data selection, data cleansing, and data pre-processing, to perform classification, clustering and knowledge representation. As an emerging area, a lot of research can be performed in this field. The main areas of research include agent-based data warehouse, agents for information retrieval, agents for distributed and parallel learning, information gathering agents and mobile agents for distributed data mining [3].

2. Agent Mining

Data mining is a multidisciplinary process of discovering interesting patterns in large data sets in order to assist the decision making process. Agent mining refers to the application of autonomous intelligent agents in the field of data mining in order to support and enhance the knowledge discovery and decision making process while providing high performance and scalability. Due to their autonomous, flexible, mobile, adaptable and rational nature, agents are an excellent choice for parallel, multisource, distributed mining. Fig. 1 shows agent mining as a two way process. It consists of agent driven data mining as well as data mining for agents. In agent driven data mining, for instance, agents can be used for data selection, data integration, data preprocessing, classification, clustering, association rules mining as well as knowledge presentation. In data mining for agents, data mining is used to extract knowledge from large datasets in the form of decision trees or data induces rules, which provide logic for the intelligent agents. For instance, consider an enterprise resource planning system that maintains a log of all decisions and actions taken by a company. Using data mining, the developer can identify, code and encapsulate the logic behind these decisions and actions into agents that are robust and trustworthy enough to replace the human decision making process.

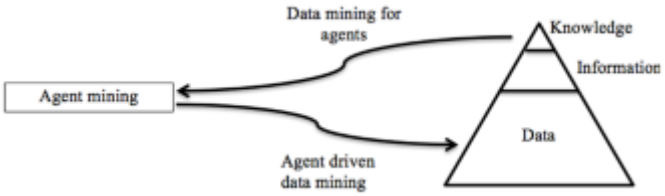

Fig. 1 Agent mining, a two way process

The agent-data mining collaboration may occur and can be analysed in a number of diverse dimensions:

- Resource dimension at data, information, and knowledge levels.
- Infrastructure dimension at infrastructure, architecture, and process levels.
- Learning dimension at learning methods, learning capabilities, and performance levels.
- Interaction dimension for coordination, cooperation, negotiation, and communication.
- Social dimension in social and organizational factors for instance, in human roles.
- Performance dimension in the performance enhancement of one end of the coupled system.
- Interface dimension at the human-system interface, user modeling and interface design level.
- Application dimension in applications and domain problems [6].

3. Agent-Based Distributed Data Mining:

A huge amount of data is stored in databases. For example, supermarkets record every purchase transaction made. Within these databases there is the potential to discover new knowledge about the world. For example, a supermarket could discover that every person who buys a lasagne for two on Saturday, also buys a bottle of red wine. This can allow promotional offers and so on to be formulated. Credit card companies may find that there are common patterns in bad repayment cases. This may lead them to augment their rules for refusing increased credit limits. With the growth of networked computing, many of these databases are now distributed over a number of computers. A number of systems have already been developed to extract this kind of knowledge from databases.However, in general they discover numeric or propositional knowledge from non-distributed data. We intend to produce a system to discover first-order knowledge from distributed databases. For example, propositional algorithms cannot discover the concept of "grandparenthood" from a database containing the names of people and their parents. However, it is possible to do so using first order learning techniques. Data mining technology has emerged as a means for identifying patterns and trends from large quantities of data.Distributed Data Mining (DDM) aims at extraction useful pattern from distributed heterogeneous data bases in order, for example, to compose them within a distributed knowledge base and use for the purposes of decision making. A lot of modern applications fall into the category of systems that need DDM supporting distributed decision making. Applications can be of different natures and from different scopes, for example, data and information fusion for situational awareness; scientific data mining in order to compose the results of diverse experiments and design a model of a phenomena, intrusion detection, analysis, prognosis and handling of natural and man-caused disaster to prevent their catastrophic development, Web mining ,etc. From practical point of view, DDM is of great concern and ultimate urgency.[1-3]

The increasing use of multi-database technology, such as computer communication The networks and distributed, federated and homogeneous multidatabase systems, has led to the development of many multi-database systems for real world applications. For decision-making, large organizations need to mine the multiple databases distributed throughout their branches. The data of a company is referred to as internal data whereas the data collected from the Internet is referred to asexternal data. Although external data assists in improving the quality of decisions, it generates a significant challenge: how to efficiently identify quality knowledge from multidatabases.

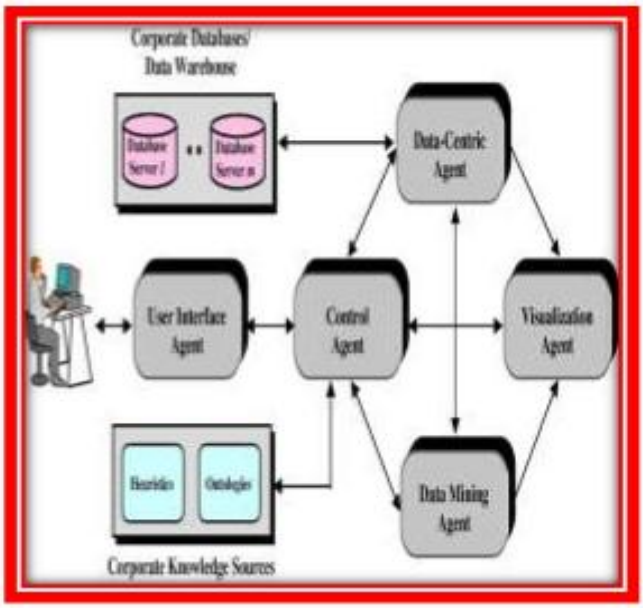

Fig. 2 Agent-Based Distributed Data Mining

4. DM Task Planning and Flow of System Operations:

DM task planning is realized by negotiation between the facilitator agent and mining agents through the message passing mechanism. Suppose a user agent sends a request to the facilitator agent to inform that it would like to do data mining with other agents in the organization. The user agent also needs to give information of model definition (dependent and independent attributes, attribute type (numeric or categorical), model type (linear or nonlinear) with its request. When the facilitator receives the request from the user agent, it negotiates with the broker agent to determine which agents to launch for this task. For example, if the user wants all possible rules meeting minimum support and confidence levels, across all available data sources, then the mining agent must ask every data agent possible for all association rules. If the user wants to find all items Y which have statistical significance for a given X, the mining agent must ask only those agents which have information about X. It would be time consuming and wasteful to ask agents which have access to data not containing X, as no rules would be generated.Finally, if the user specifies a X and a Y and asks for the level of support and confidence between the two, the agent must again only ask those agents that have information about both X and Y. The mining agent is then responsible for completing the task, while

the facilitator agent continues to plan future DM requests. When the mining agent is completed it returns the results and the facilitator agent passes them onto the user agent.

5. Open Issues and Trends

An obvious limitation of the systems discussed in the previous section is that they require a central control in which could risk the central-point-of-failure. This issue is important as the robustness could be boosted if such risk is conquered. Fortunately, peer-to-peer agent-based data mining systems are capable to effectively overcome this drawback. Moreover, the interaction and integration between the two technologies have explore the new challenges. From one side, agent computing often integrates with some other machine learning techniques as an underlying layer to cope with the problem domain, while data mining lies as a foundation of DDM systems. Various distributed computing techniques could have been employed and extended to further possibility, however maintain the ability of data mining. A novel very perspective but poorly researched application area of agents and data mining synergy is mobile, ubiquitous and peer-to-peer computing. A specific feature of such computing systems is that the latter operate with dynamic set of information sources. E.g., the mobile devices may move and freely enter to and exit from the network thus changing the set of network nodes and communication topology, changing the set of available services as well. Examples of such application areas are, e.g., smart space and ambient intelligence. In these environments, decisions are made on the basis of fusion of information received from distributed sensors and mobile devices populating the environment. One of the objectives of such application is adaptation to multiple human habits that can be achieved through learning of multiple human profiles. On the other hand, for class of applications in question, multi-agent approach supplies for most natural framework, appropriate architecture, as well as design technology. Thus, integrating agent and data mining in ubiquitous environments like smart space, ambient intelligence, etc., could be very perspective and promising to reach high quality performance of corresponding applied systems. Ongoing research is integrating existing technologies from various aspects, such as web technology, mobile devices, alternative computing technologies, databases, etc. Considering different ingredients for the integration could be a key to rapidly enhance the development process and usability of the system. Let us examine these factors.

1 Research Perspective

Data distributions in real-life applications are either homogeneous or heterogeneous. Data can be partitioned both vertically and horizontally, and furthermore data splitting may not be available across the sites. For examples, two related customer databases may not reflect each other's in which a customer may never provide contact details but somehow appear to buy some products. The applications will require a data mining technology to pay careful attention to the distributed computing, communication, and storage of the system. Another approach to develop ADDM is an inspiration from the nature which has proven to be promising. Swarm intelligence is closely related to intelligent agents. Recently, researchers pay attention to the possibility to implement DDM systems with swarm intelligence. Sample applications of swarm intelligence in data mining are rule-based classifiers using ants, feature selection with ant colony optimisation, data and text mining with hierarchical clustering ants, etc.

2 Software Engineering Perspective

Expectedly, ADDM frequently requires exchange of data mining models among the data sites. Therefore, seamless and transparent realisation of DDM technology will require standardised schemes to represent and exchange models. Therefore, in terms of software engineering, software engineering tools that support the design of data mining and distributed database would be ideal. So far, PMML, the Cross-Industry Standard Process Model for Data mining (CRISP-DM), and other related efforts are likely to be very useful.The very basic foundation of our focus is the database. Not only full-scale database, like relational database, is taken into consideration during system integration. Desktop and light-weight database running on limited devices, such as mobile phones, can be integrated into ADDM. Mobile agents can be migrated (downloaded) and perform task on the devices and take back only a representative model for further analysis.

3 System Perspective

When we focus at the system side, we find that centralised processing has a flaw in central-point-of-failure. Recently, peer-to-peer (P2P) computing has proven its excellence through its product, such as peer download software, file sharing software, which they gather users to join the service quickly. P2P is respected as one of the best scalable system, and thus it increases availability of the system as millions of peers can be attached to the network. P2P algorithm does not rely on a central server, each unit performs its own task and requests for data from others if available in order to save the redundant time. However, security is a critical issue in P2P due to exchanging information with other peers that can add a vulnerability to the network, such as denial of service or selfish behaviour. Some peers may only consume others' resources while they do not provide to others. Nevertheless, each peer must agree of terms and conditions of use before joining the network. P2P has caught researchers attention due to the compliance with multi-agent systems as appear in [19] and [11]. The last but not least item to mention is Ubiquitous computing (ubicomp). Ubicomp is a post-desktop model of human-computer interaction in which information processing has been thoroughly integrated into everyday objects and activities. Ubicomp concepts are small, inexpensive, robust, and networked processing devices. It distributes at all scales throughout everyday life. For example, a ubiquitous devices that attach to a human body to monitor his health condition can be determined as a source of data. In this way, Ubicomp can be viewed as a DDM system and is very promising as it interacts closely with human, therefore chance for collecting data is considerably high. In fact, ubiquitous and mobile computing form a novel and very perspective, although poorly researched, application area of agents and data mining synergy. A specific feature of such computing systems is that the latter often has to handle with dynamic set of information sources. E.g., the mobile devices may move and freely enter to and exit from the network thus changing the set of network nodes and communication topology, changing the set of available services as well. Examplesof such application areas

are, e.g., smart space and ambient intelligence. In these environments, decisions are made on the basis of fusion of information received from distributed sensors and mobile devices populating the environment. One of the objectives of such application is adaptation to multiple human habits that can be achieved through learning of multiple human profiles. On the other hand, for class of applications in question, multi-agent approach supplies for most natural framework, appropriate architecture, as well as sound design technology. Thus, integrating agent and data mining in ubiquitous environments like smart space, ambient intelligence, etc., could be very perspective and promising to reach high quality performance of corresponding applied systems. [15] presents a summary of challenges integrating ubicomp with MAS for data mining task.

4 User Perspective

Finally human-computer interaction issues in DDM offers some unique challenges. It requires system-level support for group interaction, collaborative problem solving, development of alternate interfaces (particularly for mobile devices), and dealing with security issues.

6. Conclusion:

Agent and distributed data mining interaction and integration has emerged as a prominent and promising area in recent years. The dialogue between agent technology and data mining can not only handle issues that are hardly coped with in each of the interacted parties, but can also result in innovative and super-intelligent techniques and symbionts much beyond the individual communities.The multi-agent systems can also be optimized for a better performance using some of the techniques described in this paper, thus further increasing their efficiency. Thus in conclusion we state that the concept of agent mining and the multi-agent systems have gained a huge momentum in the recent years and have a capability of delivering far more. More research in this respect can develop the data mining systems to a greater extent making them more efficient and increasing the accuracy of the mined data.

References

[1] J. Han, M. Kamber and J. Pei, Data Mining Concepts and Techniques, 3rd ed., Waltham, USA: Morgan Kaufmann, 2012.

[2] A. Fariz, J. Abouchabaka, N. Rafalia, “Using Multi-Agents Systems in Distributed Data Mining: A Survey,” Journal of Theoretical and Applied Information Technology, vol. 73 No. 3, pp. 427-440, March 2015.

[3] L. Cao, G. Weiss, and P. S. Yu, “A brief introduction to agent mining,” Springer, May 2012.

[4] S. J. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 2nd ed., Upper Saddle River, New Jersey, USA: Pearson Education Inc., 2010.

[5] N. R. Jennings and M. Wooldridge, “Applications of Intelligent Agents,” Springer-Verlag: New York, USA, 1998.

[6] L. Cao, V. Gorodetsky, P. A. Mitkas, “Agent Mining: The Synergy of Agents and Data Mining,” IEEE Intelligent Systems, pp. 64-72, May/June 2009.

[7] A. M. Al-Barky and J. Ali, “Intelligent mining agent,” in 8th International Conference on Computing Technology and Information Management (ICCM), 2012, vol. 1, p. 23.

[8] E. Serrano, M. Rovatsos and J. A. Botía, “Data mining agent conversations: A qualitative approach to multiagent systems analysis,” Journal of Theoretical and Applied Information Technology, vol. 73 No. 3, pp. 132-146, March 2015.

[9] O. Kazik, K. Pešková, M. Pilát and R. Neruda, “Meta learning in multi-agent systems for data mining,” pp. 433-434, Elsevier, Jan 2013.

[10] D. Sharma and F. Shadabi, “Multi-Agents Based Data Mining for Intelligent Decision Support Systems,” in 2nd International Conference on Systems and Informatics (ICSAI), 2014, pp. 241-245.

[11] Q. Li and R. Khosla, “Performance Optimization of Data Mining Applications Using a Multi-layered Multi-agent Data Mining Architecture,” in IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA), 2005, pp. 227-231.

[12] Z. Geng and Q. Zhu, “A Multi-Agent Method for Parallel Mining Based on Rough Sets,” in Proceedings of the 6th World Congress on Intelligent Control and Automation), June 2006. pp. 5977-5980.

[13]. V. Gorodetskiy. Interaction of agents and data mining in ubiquitous environment. In Proceedings of the 2008 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT’08), 2008.

[14]. V. Gorodetsky, O. Karsaev, and V. Samoilov. Infrastructural Issues for Agent-Based Distributed Learning. In Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology, pages 3–6. IEEE Computer SocietyWashington, DC, USA, 2006.

[15]. V. Gorodetsky, O. Karsaev, V. Samoylov, and S. Serebryakov. P2P Agent Platform: Implementation and Testing. In Proceedings International Workshop ”Agent and Peer-toPeer Computing”(AP2PC-2007) associated with AAMAS-07. Honolulu, Hawaii, pages 21–32, 2007.

